Data Collection and Preprocessing Phase

Date	25 September 2024
Team ID	SWTID1726888137
Project Title	intelligent handwritten digit identification system for computer applications
Maximum Marks	2 Marks

Data Quality Report

A **Data Quality Report** for an intelligent handwritten digit identification system is a structured document detailing the quality of data used in training and evaluating the model. This report assesses the accuracy, completeness, and usability of the data to ensure that the system's predictions are reliable and effective across various real-world applications.

Data Source	Data Quality Issue	Severity	Resolution Plan
Dataset (MNIST)	Dataset may lack variety in handwriting styles, leading to limited generalization for diverse handwriting.	Moderate	Augment data with handwritten samples from different sources or use data augmentation (rotation, zoom, shift, etc.) to improve model robustness to diverse handwriting styles.
Dataset (Real-time)	Captured images from the drawing canvas might have incomplete digits or partial strokes.	High	Implement a check on the drawn image to ensure minimum completeness and prompt the

			user if the drawing appears incomplete.
Dataset (Real-time)	Noise or background clutter in image captures from the canvas may affect the model's prediction accuracy.	Moderate	Apply denoising techniques or thresholding (e.g., Otsu's thresholding) to enhance digit clarity, and filter out background noise.
Dataset (Both MNIST and Real-time)	Inconsistent image scaling or padding, causing size variations that might impact model performance.	Moderate	Normalize all images to a consistent size and apply padding or resizing before inputting them to the model.